

An Efficient Linear Text Segmentation Algorithm Using Hierarchical Agglomerative Clustering

吳智瑋, 曾秋蓉, 蔡文能

Computer Science & Information Engineering

Computer Science and Informatics

judyCRT@chu.edu.tw

Abstract

Linear text segmentation aims at dividing a long text into several topical segments. It is beneficial to many natural language processing tasks, such as information retrieval and document summarization. In this article, an efficient linear text segmentation algorithm based on hierarchical agglomerative clustering is presented. The proposed linear text segmentation algorithm is implemented without auxiliary knowledge base, parameter setting, and user involvement. Experimental results show that the proposed linear text segmentation algorithm not only provides linear time computational complexity, but also provides comparable segmentation accuracy with several wellknown linear text segmentation algorithms.

Keyword : segmentation, hierarchical agglomerative clustering, computational intelligence, NLP application